



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

THEORY OF STATISTICAL TABULATION.

By G. P. WATKINS.

This brief paper deals with statistical tables in their most general aspects and is therefore labelled the "theory of tabulation." But it is a product of experience and indeed was conceived and in part written as a general introduction to directions and rules of tabulation for use in a statistical office. Hence in form it consists largely of statements of how things should be done. But the purpose and function of statistical tables are the fundamental thoughts throughout.

Nature of Tabulation. The general meaning of the word "table" appears to be an even flat surface with breadth not disproportionately small in comparison with length or, concretely, an object characterized by the possession of such a surface. The arrangement of ordinary reading matter is in a line or lines, while a statistical table presents itself as a surface.

The table thus differs from the ordinary page of letter type not merely in being composed mainly of figures, but also in being readable in two dimensions, that is, at least vertically as well as horizontally. "Reading matter" may also be a list of numbers. But the arrangement of the line (or "lines") of ordinary reading matter running back and forth on the page is not on a surface plan. A line of running print can be followed but one way. Such a line is like a string of beads, but with the type (as the beads) interrupted on the parts of the string extending from right to left and in position on the string as the line passes from left to right. The reader's eye must follow the string. A statistical table, on the other hand, can be read either down or across. It utilizes the dimensions of a surface. According to this conception, a list is not a table and a single column does not constitute a table.

A table may also sometimes be read diagonally, especially one of content and form such as to show correlation. The ages of men and of their wives, the age and the grade of school children, etc., may conveniently be compared with reference to the most frequent combinations in this way.

Matter not of a statistical character may also be put into a table when there is some advantage in reading it more than one way. Numerical data, whether statistical in character or not, are frequently best so arranged. The tabular form is used to furnish data for, and facilitate the processes of, computation, as in the familiar tables of logarithms, trigonometric functions, roots and powers, etc., and in interest tables. Here compactness of form and ease of reference are the important considerations, but these are also the reasons for being of the statistical table.

The implied division of numerical tables into two species, mathematical and statistical, suggests a question as to what is the difference between the two. The answer is that the first species contains abstract numbers, and the second, numbers that are at least relatively concrete. Statistical tables consist of numbers representing quantities or degrees of *concrete* things, qualities, or events. Hence the importance of statistical units and of their definite and constant significance. Indeed, the writer would describe statistics in general as concerned with concrete numbers and quantities and their relations. It constitutes a characteristic method or methods of dealing with such numbers, and also consists of the material appropriately so dealt with. These two aspects of the subject tend to be recognized in ordinary speech by the use of a singular verb with "statistics" in the first sense, while in the second sense the word is treated as plural. For statistics, in either sense of the word, the importance of the table is evident. This conception of statistics, it may be added, has important general bearings not involved with its incidental use in the present connection.

Tabular presentation has conspicuous advantages as regards economy of space and of time: of space, wherever the same class designation or name is to be applied to a large number of items brought together in the table in a single line or a single column; of time, on the part of those seeking information on a specific point, in that, by using line and column as guides, the specific fact sought can be found directly. These uses of the tabular form are not peculiar to numerical tables.

Tabulation, like speech, is a device for expressing ideas, and in particular for expressing them compactly and in a way to facilitate comparison and show relations. Ordinary linguistic symbols, arabic and other numerical notation (including the symbolic use of position), rulings and spacial relations, and sometimes forms special to tabular notation, are all employed for this purpose. As with language generally, the tabular presentation of facts should say as much as possible with a meaning as unmistakable as possible in as small a compass as possible. There should be no ambiguity, hence, for example, blanks should mean but one thing. Expression should be as direct as possible, hence, for example, information essential to a prompt grasping of the meaning of the table should not be put in footnotes if avoidable. Reasonable conventions regarding the use of symbols should be observed.

The table is the fundamental means of presenting statistical material and is so characteristic of the method that it may be considered the matrix of statistics. Those who first gave to "statistics" its present meaning, as distinguished from its older sense of "political" science, were opprobriously called "*Tabellenknechte*." As early as the '40's of the nineteenth century, New York State provided for the publication of railroad reports *in tabular form*.* Statistical competence may well be described as knowledge and skill in making and interpreting tables of concrete numerical data.

Uses of a Statistical Table. The stub of a statistical table is most commonly a geographical classification. For groups of such classes there will usually be sub-totals which condense the more detailed classification. But the stub may consist of the names of reporting entities, as in the case of many primary tables of corporation and financial statistics. The most important statistical data for public-service corporations are usually printed in such form by the various supervising commissions, including the Interstate Commerce Commission. But for much such data, especially for the distinctively statistical as opposed to the financial part, the company unit has little significance and compilations are made by geographical or other groups of companies. Where the facts are presented

*1907 Annual Report of the N. Y. Public Service Commission for the First District, Vol. 1, p. 452.

by reporting entities, the tabular form may serve the purpose merely of saving space, but the totals, which are of more statistical interest, are best obtained, and their composition best shown, by way of a table. If it were possible to provide the necessary space, it would of course be best always to tabulate by such return or report units, so that the person who used the primary data could make his own groupings and combinations. However, especially where the enumeration or report unit is the individual or the private family, aggregate presentation is unavoidable. Hence the stub-items of a table represent classes, rarely also composite individuals. In publishing statistics of manufacturers and other private business enterprises, the presentation of the facts for one or few companies by themselves is expressly avoided as tending to reveal the operations of individual establishments to competitors. Such procedure on the part of the U. S. Census Bureau and the various bureaus of labor statistics is undoubtedly wise administratively, though the fact that a large business corporation with stock broadly owned cannot properly withhold from the public any sort of statistical or financial data that is of general interest should be recognized and doubtless will in time be accepted in practice. But at present only quasi-public corporations appear to be dealt with statistically according to this principle.

The statistical interest of a geographical stub is, of course, not of the highest rank. The consideration determining its use is the fact that a general or primary table is in the first instance a record and repository of data. Only to a very subordinate extent is it wise to attempt to exhibit relations and significance in such a table. In a derivative (analytical or text) table the interest is of course different. But the arrangement of the items even of a geographical stub may be made to serve the purpose of explanation where, for example, the order of magnitude or of density is followed. In the New York First District Public Service Commission reports, the arrangement of lighting companies within groups determined by intercorporate relations in the order of size (amount of revenues) somewhat increases the statistical interest of the stub, since it is a step towards making the table show correlation.

It also puts first the companies in which a reader is likely to be chiefly interested, thus facilitating reference—which fact is doubtless of more practical importance than the slight aid afforded to interpretation. The order of the street-railway groups of companies in the same series of reports is in a general way that of expensiveness of line construction. These touches of correlational arrangement are suggestive of a use of tabulation which seldom affects primary tables. The correlational use, however, supposes the captions as well as the stub-items arranged according to the degree of some quality, and thus it involves cross-classification. Primary tables ought to be planned with reference to such possible use. Perhaps the presentation of such cross-classifications might well take the place of some geographical detail.

A statistical table is often merely, and always incidentally, a presentation of items going to make up a total or series of totals. The separate columns may accordingly contain things having little or no relation to each other and they may be given together merely to save space by making unnecessary the repetition of the stub. The unity of a table, however, will usually mean more than this. But it is doubtless the first or simplest purpose of a table to show this or that aggregate and how it is made up. The stub-items constitute the individual or class names for the things of which the numbers are the entries. The entries are themselves usually aggregates. But it is possible to use the tabular form for a mere tally sheet, in which case the entries represent the individual things.

In general the stub-item of a statistical table stands for a group or class of things, and the stub contains the terms of a classification. Classifications in statistics, it should be noted, must be comprehensive, hence there is usually need of an "other" or "miscellaneous" class, and commonly also of an "unknown" or "not specified" class. For the rest, all the principles conducive to right classification apply to stub and caption classifications.

It is above implied that the captions, also, as well as the stub-items, will usually constitute a classification, or perhaps more than one classification. The fact that columns commonly

add across to a total column supposes this situation. The statistical table thus becomes a mode of cross-classification.

In this more highly evolved use of the tabular form, a statistical table is essentially an arrangement of numerical data by which the data are cross-classified according to two sets of terms, those of the stub and those of the captions. The device of sub-classification is also frequently introduced in the captions and stub by way of compound captions, sub-division of stub-items, and sub-totals. The more complicated classifications usually require additional tables in series.

Instead of the terms of a classification, a time series, especially a succession of years, may be used in the stub and have much the same relation to the entries, except that column totals are then not always significant. But such a table is usually derivative.

Limitations upon Tabular Presentation. Cross-classification corresponds to what is known in algebra as combination and is covered under the topic, "Permutations and Combinations." The mathematical principle is that the number of possible different combinations of one set of things or classes of things (enumerated in the stub-items, let us say) with another set (enumerated and described in the captions) is equal to the product of the number of items in each set. This gives the number of cross-classes or entry-places in the table. There should be occasion to use most of these, or else the form of the table needs revision, or at least condensation.

The fact that cross-classification is a process of combination serves to bring out an important limitation upon the possibilities of tabular presentation. It is often desirable to show the associations or combinations of the units under three classifications or sets of cases. If the third of these classifications is merely twofold, the space required is merely double what it was before. If there are 12 rubrics under the third classification, the normal requirement is for 12 times as much space, or probably 13 times as much, since a total of the 12 classes will be desirable. If the original stub provides for 30 items and there are 10 columns, a presentation of all the possible combinations with a further series of 12 classes will require $30 \times 10 \times 12$, or 3,600 cross-classes or entry-places.

If it is desired to show completely by tabulation the relations between nativity in 12 classes, age in 10 classes, sex in 2 classes, residence in 50 classes, and occupation in 100 classes, supposing every possible combination will require an entry-place, the number of cross-classes will be $12 \times 10 \times 2 \times 50 \times 100$, or 1,200,000. If the 50 residence rubrics are made the items of the stub and 10 columns may be put on a page, that would mean 500 entry places to a page. The presentation of the facts would, therefore, require 2,400 pages. But the number of rubrics under each classification is fewer than it might be desirable to use. The above computation, moreover, does not provide for totals. Of course, much space could in practice be saved by reason of the omission of provision for impossible or infrequent combinations. Young children, for example, will not be found in occupations. However, the limitations upon what we may call *complete tabulation* are evident. The size of census volumes, even with their limitations, is thus explained.

The difficulty in question is avoided by seldom attempting complete tabulation. Some of the combinations are not important or not of special interest. The classification of those in a specific occupation by nativity, for example, is of interest for comparatively few occupations and comparatively few localities. It may often be assumed that the variation within one kind of classification in terms of another classification will be so small that a presentation of the facts for all of the first class combined will sufficiently meet ordinary statistical requirements. Detailed compilations also may often be made to serve for a number of years, provided the proportions found are representative and quite constant. The frequent necessity of resorting to such methods—the necessity in particular of using alternative classification instead of cross-classification—explains why a given statistical compilation will seldom enable one to answer all the questions for which a solution is sought. The facts are contained in the returns but they cannot all be presented.*

*Table XI of the 1911 street-railway report (in the volume on transportation statistics, Volume II of the 1911 Report of the New York Public Service Commission for the First District), dealing with Accidents, shows in Division C a classification of injuries by occasion and a separate classification by degree of seriousness, but the relations between the two classifications are not shown, that is, the classifications are

A report schedule from which tabulations are made is commonly itself in tabular form and may contain a cross-classification. Only one who has had practical experience with the problem of devising a general table or tables to contain what is most important in such returns can appreciate the difficulty of obtaining satisfactory results in a limited space. But the reader is prepared for an application of the theory of mathematical combinations to such a case. If only 50 such report schedules are to be tabulated in a way to show the individual returns and supposing the schedule has 10 stub-items and 20 captions, then in order to present *all* the facts it would be necessary to provide at least 200 columns of 50-line tabular matter. Alternative tabulation, on the other hand, which would utilize only the cross and down totals of the schedule, would require 30 columns. It is assumed, of course, that the data of each schedule are themselves aggregates and that each such aggregation has interest of its own. If only the totals for the 50 returns taken together are wanted, only as many entry-places are required as are contained on one of the schedules, that is, $20 \times 10 + 31$ (for totals), or 231 in all—which is a table of modest dimensions. Enumeration schedules, it should be noted, are not often of a character to raise this question in just this form.

Detailed classification according to geography or locality is, as has been stated, not of statistical interest in proportion to the amount of space it takes in primary statistical publications. Every locality, however, has a neighborhood interest

alternative and do not make a cross-classification. Disregarding the difference between "accidents," "killed," and "injured," under occasions—which in fact presents in part the facts regarding the seriousness of the result—there are 5 occasion rubrics (as condensed from 22 in the annual report form) and 6 seriousness rubrics. To present all the possible combinations of these would require 30 columns instead of the 11 at present required. The same facts, with a somewhat more detailed classification by occasion of injury (disregarding, however, the number of accidents) are sub-divided between passengers, employees, and others in the three parts of Division D. If, instead of the mere distinction between killed and injured in Division A of this table, the subdivision provided for all classes of injuries and a total, 4×11 , or 44 columns, would have to be added. Six columns might be dispensed with, but should be kept as totals.

A reduction of what would otherwise be the undue length of Tables XXXIII and XXXIV in the employees and wages statistics of the 1911 report for lighting companies (Volume III of the 1911 Report of the New York Public Service Commission for the First District) is effected by using a condensed stub in which systems or groups take the place usually occupied by individual companies. This requires the preliminary tabulation of company returns to get the totals thus printed. The use of the full company stub would increase the length of Table XXXIII in the ratio of approximately 8 to 36. Instead of occupying 19 pages, it would take 86.

in facts about itself, which the Census Bureau and other statistical officers feel called upon to cater to. Statisticians, also, often appear to want a great amount of local sub-division of the primary data.* This demand is largely the result of attempting to show the degree of connection between various sorts of social conditions by way of comparative cartograms or of corresponding numerical analysis, a method which is in effect crudely correlational. This purpose could be better served by re-counts of the punched cards pertaining (say) to a given city, with reference not merely to showing the relation between certain conditions and certain localities, but to tracing actual connections (the individual or family being the unit) so far as the data compared came from the same schedules. But for comparison with data from diverse sources the cruder cartogrammatic method might be necessary, which, however, would be aided by an adaptable locality classification. For such purposes the re-counting of census cards by responsible private agencies, as well as by the Census Bureau itself after the decennial rush is over, ought to be facilitated and encouraged wherever it would serve a public object. Something of this sort, rather than further local detail, is the true statistical desideratum.

One way in which the Thirteenth Census meets the problem of voluminousness resulting from geographical details is interesting in this connection. Instead of such details being furnished for all the states together, the Abstract has a supplement for each state giving the geographical detail for it alone. Thus such local details are furnished only so far as they are interesting and useful to each class of readers.

With our present-day mechanical facilities for "tabulation," the process of sub-division and cross-classification of aggregates is limited rather by the degree of significance of the results, and by the cost and awkwardness of voluminous reports, than by the time required to make the necessary sortings and counts of cards already punched. While the mathematical theory of combination is a good point of departure in planning

* Cf. Robert A. Woods, Unit Accounting in Social Work, *QUARTERLY PUBLICATIONS OF THE AMERICAN STATISTICAL ASSOCIATION*, March, 1913, Vol. XIII, p. 361. This paper was read at the 1912 annual meeting. Unfortunately, the discussion that followed is not printed.

tables, most combinations of the terms of diverse classifications, even if they occur, have no concrete significance.

Comprehensiveness, Comparability, and Compactness as Essentials of Good Statistical Tables. The significance of a statistical table, as of statistics generally, depends very largely upon its being comprehensive for the field it covers. Truth in its statistical aspect is representativeness. The only absolute guaranty of the representative quality of an aggregate is that it reflects all the units within its scope. According to the mathematical theory of probabilities, much less is necessary, but this theory does not take account of the selective tendency of events and of observation, for which the statistician must be continually on his guard. The point is illustrated by the well-known difference in quality between results obtained by complete enumeration and those obtained from a circular letter or questionnaire.

A table should not be composed of mere samples. It is better to make it of narrow scope but comprehensive as far as it goes, *i. e.*, within its territorial or other limits. A table, furthermore, is likely to be one of a series, which should all be on the same basis or, at least, conform sufficiently to the basis of the series so that its representative quality and the comparability of its totals are not appreciably impaired. The most surely understood uniform basis, meeting all the requirements of comparability, is the comprehensive basis. When a table falls short of the basis of its fellows, but in a way not such as to compel its omission altogether, the appropriate place to indicate what is lacking is a general note. Sometimes it may be well to have two sets of totals to a table, one on the most comprehensive basis, and one less comprehensive, but such as to supply aggregates for data that, though falling short of perfect comprehensiveness, may be of qualified value in other ways, as for example, in the computing of ratios. On the other hand, if it is desirable to present information in connection with only one of a series of tables, it is well, in order to avoid impairing the comparability of one table with the others of the series, to put the data that exceed the standard scope in brackets and not take them into the totals, thus letting them be *in* the table for purposes of

reference, but not strictly *of* it. Uniform comprehensiveness upon some definable basis is the ideal standard. Even a small per cent. impairment of comprehensiveness may mean a large decrease in tabular efficiency.

The same principle applies with reference to corresponding tables for a series of years. While it is desirable that new data be made use of, full notice of a change of basis should be given and it is often well to give figures and make comparisons on both the old and the new basis for the first year of the change. Especially in derivative tables attention to comparability is imperative, without regard to cost in the way of added complexity, etc. Ratios, for example, should usually be given on both bases where there is a change. This again is a question of representativeness, though here differences between aggregates, rather than the aggregates themselves, are under consideration. How important this question is in another of its phases is illustrated by the place commonly given to averages, *i. e.*, representative numbers, as the gist, if not the substance, of statistics.

The complement of the requirement of comprehensiveness is that of compactness. It is of the essence of a table to convey a large amount of information in a small space. Hence sparsely tenanted columns are an eyesore, and blank columns, even where the original classification may have reasonably planned to use them, should not be tolerated. Blank lines are hardly less justifiable. Classifications should be revised when the data as spread out show such waste of space. Unrepresented classes may be disposed of in the notes. Sparsely tenanted columns should be consolidated, subdivisions of entries being indicated by footnotes if desirable. A "miscellaneous" column may often be employed with reference to such residual classes. It should never include more than a small per cent. of the material of the table. But sometimes the desirability of keeping up tables on a uniform plan, *e. g.*, through a series of years, may justify continuing sparse columns till a comprehensive overhauling of the form of tables is undertaken.

The table must ordinarily be planned with reference to fitting the printed page, as single-page lengthwise, single-

page upright, twin upright, or as a series of such. Hence dimensions in terms of columns and lines must often be carefully studied before being finally fixed. The large page and the resulting unwieldy size of most statistical volumes are due to the need of space for manoeuvering the tabular matter. Often the presentation in sections of what is functionally one table becomes necessary.

General Tables and Derivatives Tables Distinguished. A table serving primarily the purpose of a repository of comprehensive statistical data is distinguished as a general table, also, with reference to its being closest to the original data, as a primary table.

Derivative tables are summaries and auxiliary ratio tables. They may usually be distinguished as text or analysis tables. But some ratio tables, or at least some ratios, are often included among general tables. Derivative tables are based upon general tables and contain matter suitable for incorporation in analysis. They may vary in form from year to year according to the exigencies of the situation and according to the points emphasized in the text. Unlike the general tables they will usually contain data and comparisons, including absolute and per cent. increases, for several years. Just as general tables serve to show in terms of absolute numbers the composition of aggregates, a derivative table frequently serves the purposes of explanation correspondingly by means of per cent. distribution. If text tables contain data taken direct from returns, these are so treated because of lack of comprehensiveness in the data, or of perennial interest in that kind of data. Explanatory and qualifying statements contained in general-table footnotes should, unless unimportant, be either repeated or referred to in footnotes, or in text immediately adjacent to the text tables.

It is the common practice of statistical bureaus to number tables serially for each report. If Roman numerals are used for the general tables, arabic numerals are used for derivative tables, or *vice versa*. The United States Census has in general employed arabic numerals for the serial numbers of general tables and roman numerals for text tables, but in the Thir-

teenth Census volumes the text-table numbers are arabic and roman numerals seem to be reserved for general tables.

No strict line can be, or need be, drawn between what should go into general and what into text tables, though the fact that ratios are logically a part of the analysis gives the analytical text, if there is any such, a strong claim upon them. Grand totals certainly go with the general tables not only as closing them up but also because of their importance as a proof check. But divisional totals serving the purpose of a summary may go in either place. Ratios, too, may come to have so thoroughly well-established a place as to be in effect a part of the data that the public will expect to find in connection with the general tables. A derivative table in a report containing the corresponding primary tables is seldom to be considered a thing by itself to the extent of requiring no reference to its sources on the part of a reader who uses it carefully.

Comparisons with previous years—or with corresponding months (or other portions) of previous years—are also strictly a part of analysis, but their significance is so direct and their meaning in general so unmistakable that some of them may well be looked for in the general tables. They are made much of especially in commercial and financial statistics. The United States Census is liberal in presenting comparisons for previous decennial years in its general tables.

General or primary tables rightly occupy the largest place in most government statistical publications. Indeed, some official statisticians feel that the preparation and presentation of the primary tables is their whole duty. But some working-over of the raw material by those directly concerned with its compilation is desirable, if for no other reason than the beneficial reaction on the original data and tables consequent upon analyzing and applying them to the solution of scientific and practical problems. Proper emphasis upon the function of such statistical publications as sources does not preclude brief suggestive analysis, in addition to the necessary descriptive and cautionary remarks.

The Rounding and Abbreviation of Numbers. The use of rounded or cut-off numbers should seldom be adopted in general or primary tables, though doubtless desirable in

derivative or interpretative tables. The practice is often recommended without reference to, or due emphasis upon, this very necessary qualification.

Even in derivative tables, the giving of a large number, for example, millions of inhabitants, to the last digit would mislead by its supposed suggestion of "spurious accuracy" only in the case of a reader who would have at least equal difficulty in understanding what the rounding of the figures meant. The notion that we should print numbers showing the digits only in so far as they are known to be accurate, or on the basis of the theory of probabilities considered to be so, is impractical to the height of absurdity. The truth of the stated population of New York City—4,766,883 in 1910—is not of a nature to imply that the figure 3 in the units place has statistical significance. The statistician knows that the last four digits are neither more nor less accurate or truthful if made to read 7,000 instead of 6,883. He does not need to be reminded that the 117 has no objective or exact meaning in such an aggregate. It is seldom necessary to indicate that large numerical aggregates are approximate as to the right-hand figures.

But there is also a positive objection to the rounding of such numbers. From the point of view of statistical administration it is important that, for example, the population of a large area be the total for all its parts down to the smallest district for which separate figures are given, some of which in the instance referred to actually have less than 117 inhabitants. Rounding an absolute number is never obligatory and should never be done in a way to deprive anyone of the possibility of completely checking the number and of using for this purpose, if for no other, the unmodified original aggregate. Primary numerical data should not be rounded.

As regards ratios, too, their mechanical computation with equal ease to a larger as to a smaller number of places makes the decision of how far they should be carried a question of conventional expectations and of economy of attention rather than anything more fundamental. This statement does not refer to (and does not apply for) slide-rule computations. The carrying out of ratios to two decimal places (or for per cent. to hundredths of one per cent.) seems to be the most

satisfactory practice for most cases, so far as fractions are desirable, though only the first place will usually be itself significant, the second serving rather to qualify the first. Where three decimal places are used, the printer, and sometimes the reader, will easily mistake the point for a comma.

But much depends on how far it is the statistician's aim to make his material popular—an end that is, of course, entirely worthy in itself. The desirability of rounded and abbreviated numbers, also of the use of few numbers, in statistical exposition is chiefly of the same nature as are the claims of stylistic elegance or of force (as a writer may prefer or the conditions require) in the use of the English language. The first duty of one presenting statistical results is to be adequate and accurate; if possible it is well for him to be also elegant, or forcible, or whatever else may be desirable, in his choice of words and of numerical expressions.

The process of rounding or cutting-off numbers is by no means simple or a matter of course. On the contrary, it requires considerable statistical technique—else totals will be found not to check with items and ratios not with the data from which they are derived. It may be noted incidentally that where it may seem desirable, as frequently in the case of estimates, to round or abbreviate both a relative number and the corresponding absolute number, one cannot do both and at the same time preserve the requisite verifiable relation between the two. This fact counts against the rounding even of estimates, though some sign of approximation is in such cases especially desirable.

Tabular Notation. The rounding and abbreviation of numbers is strictly a part of the subject of tabular notation, but so fundamental as to affect the character of the statistical table as such. The word "notation" properly refers to the relation between the signs and symbols used to convey the meaning of any part of the table and the significance arbitrarily or conventionally attaching to them. To illustrate, it would seem that the last two digits, 83, of the figure for the population of New York City in 1910, preceded as they are by five other digits having the significance of position proper to them according to the arabic numerical notation, ought, without

difficulty, to be interpreted as having a different statistical significance from the figure 83 as arrived at, for example, by a careful housewife on inventorying her pieces of silverware preparatory to putting them into safe deposit, or by a dairy-man counting his stock.

The signs used in tabulation are chiefly arabic numerals and the letters of the alphabet in their various appropriate combinations. The position of such a sign may be a part of the notation. The notation of a table is the language in which its import is expressed; and that language should be as direct, concise, and unambiguous as it is possible to make it.

The technique of statistical notation has not reached a high stage of development. The writer, at any rate, feels that the tendency among statisticians to treat a table as a mere repository of numbers and to indicate in footnotes any state of facts not so represented is objectionable. The absence of a report, the failure to segregate returns, the character of an entry as estimated or as incomplete—all these are matters that can be shown by appropriate signs on the face of the table. The best policy would seem to be to make the tabular entries self-explanatory to as high a degree as possible, for the purposes of the particular tabulation, by the use of word or other non-numerical sign entries where feasible. Footnotes are thus reserved to supplement or qualify both numerical and sign entries and especially are not intended to take the place of lacking numbers. But the technique of tabular notation lies outside the scope of a discussion of the general aspects of statistical tabulation.